

Desenvolvimento de um vocabulário controlado para o repositório institucional CarpedIEN

Development of a controlled vocabulary to the institutional repository CarpedIEN

Desarrollo de un vocabulario controlado para el repositorio institucional CarpedIEN

Luana Farias Sales | lsales@ien.gov.br

Comissão Nacional de Energia Nuclear, Instituto de Engenharia Nuclear. Rio de Janeiro, Brasil.

Lucas de Lima Rocha | lucaslrocha@gmail.com

Comissão Nacional de Energia Nuclear, Instituto de Engenharia Nuclear. Rio de Janeiro, Brasil.

Márcia Teixeira Cavalcanti | marciacavalcanti@gmail.com

Faculdades Integradas Maria Thereza. Rio de Janeiro, Brasil.

Resumo

Durante o ano de 2002, o Instituto de Engenharia Nuclear (IEN) apresentou a exposição itinerante “IEN – 40 anos” que, em 2003, chegou ao Museu de Astronomia e Ciências Afins (MAST). Parte da exposição trouxe um conjunto de fotografias que conta a história do IEN e, quando essa coleção retornou ao Instituto, decidiu-se que a guarda dela seria direcionada aos cuidados da biblioteca. Com isso, foi elaborado um projeto institucional que teve como estratégia principal digitalizá-las e disponibilizá-las no repositório institucional do IEN. Para que a recuperação dessas imagens pudesse ser mais consistente, identificou-se a necessidade da construção de um vocabulário controlado para indexá-las, e o projeto se estendeu a todos os itens do repositório. Essa proposta busca apresentar as etapas de construção desse vocabulário, desde a identificação da problemática, passando pelas estratégias de escolha das palavras-chave, a abordagem metodológica e o status atual do projeto, que ainda está em construção.

Palavras-chave: Vocabulário controlado; Curadoria digital; Memória institucional; Repositório Institucional; Informação nuclear.

Abstract

During 2002, the Institute of Nuclear Engineering (IEN) presented the itinerant exhibition “IEN - 40 years”, which, in 2003, arrived at the Museum of Astronomy and Related Sciences (MAST). Part of the exhibition brought a set of photographs that tells the story of the IEN and, when this collection returned to the Institute, it was decided that her guard would be directed to the care of the library. With this, an institutional project was elaborated, whose main strategy was to digitize them and make them available in the institutional repository of the IEN. In order for the retrieval of these images to be more consistent, the need to build a controlled vocabulary to index them was identified, and the project was extended to all items in the repository. This proposal seeks to present the stages of construction of this vocabulary, from the identification of the problem, through the strategies of choice of keywords, the methodological approach and the current status of the project, which is still under construction.

Keywords: Controlled vocabulary; Digital curation; Institutional memory; Institutional Repository; Nuclear information.

Resumen

Durante el año 2002, el Instituto de Ingeniería Nuclear (IEN) presentó la exposición itinerante “IEN - 40 años” que, en 2003, llegó al Museo de Astronomía y Ciencias Afines (MAST). Parte de la exposición trajo un conjunto de fotografías que cuenta la historia del IEN y, cuando esa colección regresó al Instituto, se decidió que la guardia de ella sería dirigida a los cuidados de la biblioteca. Con ello, se elaboró un proyecto institucional que tuvo como estrategia principal escanearlas y ponerlas a disposición en el repositorio institucional del IEN. Para que la recuperación de esas imágenes pudiera ser más consistente, se identificó la necesidad de la construcción de un vocabulario controlado para indexarlas, y el proyecto se extendió a todos los elementos del repositorio. Esta propuesta busca presentar las etapas de construcción de ese vocabulario, desde la identificación de la problemática, pasando por las estrategias de elección de las palabras clave, el enfoque metodológico y el status actual del proyecto, que aún está en construcción.

Palabras clave: Vocabulario controlado; Curaduría digital; Memoria institucional; Repositorio Institucional; Información nuclear.

1 Introdução

O Instituto de Engenharia Nuclear (IEN), na qualidade de uma das unidades de pesquisa da Comissão Nacional de Energia Nuclear (CNEN), está voltado para a pesquisa científica e tecnológica, para o ensino em nível de pós-graduação e para a inovação no domínio interdisciplinar das ciências nucleares. Como desdobramento de suas atividades acadêmicas e de pesquisa e desenvolvimento o IEN produz intensamente informações e dados de pesquisa em formatos digitais. Esses recursos informacionais se manifestam na forma de relatórios, teses e dissertações, artigos, multimídia, apresentações, aulas, programas de computador, simulações, ambientes virtuais e coleções de dados, que estão registrados nos mais diversos formatos e mídias.

Com o objetivo de preservar, integrar, recuperar e compartilhar em âmbito nacional os recursos informacionais da área Nuclear, uniu-se tecnologia e gestão na construção de um repositório digital que apoiasse a custódia e a preservação e evitasse a fragmentação da produção técnico-científica e também dos dados digitais de pesquisas produzidos no IEN. Atualmente em funcionamento, o Repositório Institucional CarpedIEN² conta com mais de mil objetos digitais, entre artigos científicos, comunicações em congressos, dissertações de mestrado, teses de doutorado, documentos históricos e fotografias, que se apresentam não só como um reflexo da produção de Engenharia Nuclear no país, mas também como um registro histórico do instituto em si.

Dentro do repositório institucional, os itens depositados necessitam de alguns metadados obrigatórios, tais como título, autoria, resumo e palavras-chave, e este último item é essencial para a recuperação dos objetos digitais por assunto, seja através da ferramenta de busca presente no repositório, seja por uma lista automática gerada pelo próprio sistema que conta com todas as palavras-chave indexadas no repositório. No entanto, ao verificar a lista presente no sistema, é possível se deparar com problemáticas que vão desde a presença de termos sinônimos até a presença de erros de digitação ou diferenciação de letras maiúsculas e minúsculas, que acarretam erros quando os objetos digitais precisam ser recuperados. Se, por exemplo, estiverem indexados os termos ‘Energia nuclear’, ‘Energia Nuclear’ e ‘energia nuclear’ em diferentes objetos digitais – exatamente com a mesma grafia, mas com diferenciação nas iniciais maiúsculas –, o usuário é obrigado a realizar a busca nas três diferentes grafias para encontrar os documentos; caso faça apenas uma busca, a recuperação omitirá alguns dos resultados.

Identificar a falta de consistência dessas palavras-chave foi a motivação principal para se iniciar a ideia de um vocabulário controlado que pudesse nortear a indexação dos objetos digitais dentro do repositório institucional. Inicialmente pensado para indexar um conjunto de fotografias que remetiam à história do IEN, verificou-se o potencial que esse tipo de tratamento da informação traria ao repositório de um modo geral, e o processo de desenvolvimento estendeu-se para todas as categorias de objetos depositados, e não apenas às fotografias.

Com isso, pode-se estabelecer que os objetivos do vocabulário controlado são: (1) aprimorar a consistência na recuperação informacional, eliminando erros referentes à sinonímias, homonímias e problemas como erros de digitação e diferenciação de letras maiúsculas e minúsculas; (2) nortear futuros depósitos dentro do repositório institucional, guiando os depositantes para que insiram informações consistentes que possam, posteriormente, ser recuperadas de forma eficaz; e (3) funcionar como um guia rápido para o entendimento de conceitos da área de engenharia nuclear, expressos através de termos com breves definições de seus significados.

2 Metodologia para construção do vocabulário CarpediEM

O processo para a construção do vocabulário controlado iniciou-se a partir da necessidade de indexação de um conjunto de fotografias que remetiam à história do IEN. Como grande parte da identificação das fotografias deu-se através de entrevistas com funcionários antigos do Instituto, a linguagem natural utilizada na primeira etapa carregou consigo as inconsistências inerentes à fala; com isso, tornou-se necessária uma normalização. Partindo de uma tabela elaborada com as explicações em linguagem natural para cada uma das fotografias identificadas, estabeleceu-se a extração de termos para a composição do vocabulário controlado, utilizando parte da metodologia proposta¹ por para as etapas de planejamento e levantamento do vocabulário.

Estabeleceu-se como público-alvo os servidores, pesquisadores e alunos vinculados ao Instituto de Engenharia Nuclear, bem como os interessados nos assuntos envolvendo engenharia nuclear; como função, estabeleceu-se que o vocabulário buscaria uma maior consistência na indexação de conteúdos no repositório CarpediEM, objetivando uma posterior recuperação informacional com maior qualidade; como forma de apresentação, optou-se pela estrutura alfabética, apresentando definição do termo, termos gerais, específicos e/ou associativos, além da presença de sinônimos não-preferidos e termos não-autorizados.

Foram extraídos os termos mais significativos das descrições, como nomes de instrumentos, processos e locais importantes dentro do instituto. Esses termos foram separados dentro das cinco categorias de Ranganathan (Personalidade, Matéria, Energia, Espaço e Tempo), utilizando o Método de Categorização com o objetivo de pensar o domínio e orientar a organização posterior dos termos.

A *figura 1* apresenta uma imagem da tabela de identificação das fotos.

Figura 1. Print da planilha de identificação das fotos

Nº FOTO	LOCAL/SITUAÇÃO	ANO/PERÍODO	DESCRIÇÃO/PESSOAS	referência as categorias criadas pela Valéria OBSERVAÇÃO
0249	Rejeitos		Edifício de Depósitos de Rejeitos Radioativos, Instituto de Engenharia Nuclear, Ilha do Fundão, Rio de Janeiro. Década de 1990. Foto colorida.	
0250	Laboratório de Medidas Radiológicas (Bloco N)		Edifício do Laboratório de Medidas Radiológicas (LMR)/Coordenação de Segurança Proteção Radiológica (CSPR) - primeira fase da construção, Instituto de Engenharia Nuclear, Ilha do Fundão, Rio de Janeiro. Década de 1990. Foto colorida.	
0252 e 0253	Serviço de Engenharia	12 setembro 1997	Funcionário Lino Cesar Ferreira Lagdem, sentado à mesa; funcionário Edson Mendes Guimarães, de pé, Serviço de Engenharia (civil, mecânica, elétrica), Instituto de Engenharia Nuclear, Ilha do Fundão, Rio de Janeiro. 1997. Foto colorida.	
0254	Ciclotron CV-28	19 agosto 1997	Vista frontal de Célula de Iodo 123 produzida pela Nuclep - Nuclebrás Equipamentos Pesados, Rio de Janeiro. Foto colorida.	
0255			Vista em diagonal da parte traseira e lateral de Célula de Iodo 123 produzida pela Nuclep - Nuclebrás Equipamentos Pesados, Rio de Janeiro. Foto colorida.	
0256			Vista lateral da parte superior da Célula de Iodo 123 produzida pela Nuclep - Nuclebrás Equipamentos Pesados, Rio de Janeiro. Foto colorida.	
0257			Vista frontal de Célula de Iodo 123 produzida pela Nuclep - Nuclebrás Equipamentos Pesados, Rio de Janeiro. Foto colorida.	
0258			Vista em diagonal de frente e lateral de Célula de Iodo 123 produzida pela Nuclep - Nuclebrás Equipamentos Pesados, Rio de Janeiro. Foto colorida.	
0259	Galpão Tecnológico de Sódio - GTS	12 setembro 1997	Circuito Térmico a Sódio CTS-1 (Loop de Sódio), Galpão Tecnológico de Sódio, Instituto de Engenharia Nuclear, Ilha do Fundão, Rio de Janeiro. Foto colorida.	

Fonte: Elaboração própria.

Após essa divisão de termos a partir da linguagem natural das descrições, optou-se pela busca de definições que pudessem nortear o entendimento sobre o domínio nuclear. Primeiramente, em glossários especializados na área nuclear e, quando não foram encontradas as definições nos glossários consultados, em ferramentas de busca online como o Google. O conjunto desses termos encontrados comporá parte do vocabulário controlado, atualmente em desenvolvimento, que tem por objetivo racionalizar os dados indexados no repositório institucional CarpediEM. A organização desse conjunto de termos separados dentro das categorias de Ranganathan e as explicações encontradas em glossários e em ferramentas de busca online foram organizadas através do software Excel.

A *figura 2* apresenta uma imagem da organização dos termos e das explicações encontradas nos glossários.

Figura 2. Print da planilha de identificação de termos com suas respectivas explicações

	A	B	C	D	E	F	G
1	Nº FOTO	PERSONALIDADE	MATÉRIA	ENERGIA	ESPAÇO	TEMPO	EXPLICAÇÕES DE GLOSSÁRIOS
329	368	Misturador para separação de urânio		Procedimento de teste	Laboratório de Química		
330		Decantador para separação de urânio			Instituto de Engenharia Nuclear		
331					Ilha do Fundão		
332					Rio de Janeiro		
333	369	Monitor Inteligente de Radiação, modelo MIR 7028			Instituto de Engenharia Nuclear		Monitor: Instrumento que mede o nível de radiação ionizante numa área. (Glossário Nuclear, p. 46-7).
334		Sonda Panqueca, modelo SPQ-7026			Ilha do Fundão		
335					Rio de Janeiro		
336	370				Estante de equipamentos		
337	371 a 373	Triga			Mesa de instrumentação e controle		Triga: reator de pesquisa do CDTN desenvolvido e produzido pelo IEN (anotação feita à mão)
338	374				Estante de equipamentos		
339	375 a 377b	Triga			Mesa de instrumentação e controle		
340	378				Caverna do Cíclotron		
341	379 e 380	Sistema Kipros		Construção	Cavernas dos alvos do sistema Kipros	20 jun. 1997	
342					Instituto de Engenharia Nuclear		
343					Ilha do Fundão		
344					Rio de Janeiro		
345	381 a 386			Construção	Laboratório de Dosimetria Termoluminescente	1998	Dosímetro: dispositivo que mede a dose absorvida de radiação, por exemplo um dosímetro dosimétrico, ou uma câmara de ionização. (Glossário Nuclear, p. 25).
346				Escavação	Instituto de Engenharia Nuclear		
347				Marcação	Ilha do Fundão		
348					Rio de Janeiro		
349	389	João Carlos Pereira dos Santos					
350	394			Construção	Laboratório de Dosimetria Termoluminescente	1998	
351				Instalação	Instituto de Engenharia Nuclear		
352					Ilha do Fundão		
353					Rio de Janeiro		
354	396	Placa de identificação da porta principal			Laboratório de Dosimetria Termoluminescente	1998	
355					Instituto de Engenharia Nuclear		
356					Ilha do Fundão		
357					Rio de Janeiro		
358	397 e 398				Instalações internas	1998	
359					Laboratório de Dosimetria Termoluminescente		

Fonte: Elaboração própria.

Durante esse processo, o grupo de trabalho percebeu o potencial deste vocabulário para todos os objetos digitais presentes no repositório. Com isso, optou-se pela extensão das buscas por definições, anteriormente resumidas às fotografias do Projeto Memória-IEN, a todos os itens presentes no repositório CarpediEN – a saber: artigos de periódicos, trabalhos de congressos, dissertações de mestrado e teses de doutorado dos servidores do Instituto, teses orientadas por servidores do Instituto, capítulos de livros e relatórios. Foram elencados um total de 1607 termos presentes no repositório, listados em ordem alfabética no software Excel, onde foram identificadas inconsistências referentes a diferenciação em maiúsculas e minúsculas, acentuação gráfica e erros de digitação, que serão posteriormente sanadas.

3 Considerações finais

Atualmente, o vocabulário controlado encontra-se em fase de construção, com a busca pelas definições em glossários especializados na área nuclear e em ferramentas de busca online, e com a normalização dos termos que possuem inconsistências terminológicas ou conceituais. Além disso, o grupo optou por excluir palavras que tenham baixa representação no domínio da engenharia nuclear ou que não sejam interessantes ao escopo do projeto, como termos muito genéricos e de amplo conhecimento em qualquer domínio intelectual.

O grande desafio de construir o vocabulário controlado é o de formular definições que funcionem tanto para especialistas na área quanto para alunos com menor conhecimento, para que todos possam utilizar o potencial desse instrumento em suas pesquisas. Também se mostra um desafio entender em que medida as definições encontradas se adequam ao termo pesquisado, sendo necessária a participação de especialistas na área da engenharia nuclear para sanar eventuais dúvidas sobre os conceitos explicados.

É pertinente acrescentar que o trabalho feito deverá estar sempre em atualização, com revisões periódicas dos termos e da sua persistência ao longo do tempo. Por mais que se trate de uma ciência exata, o domínio de conhecimento nuclear se encontra em movimento e expansão, evidenciando que novas descobertas são feitas e novos conhecimentos precisam ser incorporados ao universo descrito através deste vocabulário. Um dos grandes desafios futuros para esse instrumento terminológico é o de se manter atual às pesquisas geradas dentro do instituto, para que funcionem como parceiros dos pesquisadores no momento de indexação de seus trabalhos ou no momento de busca por trabalhos já depositados no repositório institucional que possam servir de base às suas pesquisas.

Referências

1. Campos MLA, Gomes HE, Motta DF. Elaboração de tesauro documentário: tutorial. [Internet] [citado em 19 abr 2017]. Disponível em: <http://www.conexaorio.com/bit/tesauro/>
2. Sales LF, Sayão LF. Repositório do Instituto de Engenharia Nuclear: curadoria digital, publicações ampliadas e gestão de pesquisa. Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação-FEBAB; Florianópolis: UFSC, 2013. p. 982-986.